AIR FORCE HUMAN RESOURCES LABORATORY

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: PRELIMINARY ANALYSIS

Frederick M. Siem
Thomas R. Carretta
Theresa A. Mercatante

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

May 1988

Interim Technical Paper for Period January 1983 - January 1987

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

DANIEL L. LEIGHTON, Lt Col, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-87-62	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b. OFFICE SYMBOL (*If applicable*) AFHRL/MOEA	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (*City, State, and ZIP Code*) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (*City, State, and ZIP Code*)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (*If applicable*) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (*City, State, and ZIP Code*) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62703 / PROJECT NO. 7719 / TASK NO. 18 / WORK UNIT ACCESSION NO. 45

11. TITLE (*Include Security Classification*)
Personality, Attitudes, and Pilot Training Performance: Preliminary Analysis

12. PERSONAL AUTHOR(S)
Siem, F.M.; Carretta, T.R.; Mercatante, T.A.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Interim	FROM Jan 83 TO Jan 87	May 1988	20

16. SUPPLEMENTARY NOTATION
Paper presented at the Third Annual South Texas Symposium on Human Factors and Ergonomics sponsored by the Alamo Chapter of the Human Factors Society, University of Texas at San Antonio, San Antonio, Texas, 8 May 1987.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
attitude assessment; individual differences; Basic Attributes Tests; personality assessment; classification; pilot selection

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

Developments in research concerning personality characteristics have led to a renewed interest in applications of individual differences measures for selection of pilot candidates. Recent research efforts have focused on selecting for positive characteristics, rather than screening out pathological traits. Another development is the use of tests in which the dimension of interest is not readily apparent to the examinee.

In the present study, five personality and attitude measures were administered to 883 USAF pilot candidates as part of an experimental test battery under consideration for operational use in pilot selection and classification. These tests were designed to assess decisiveness, risk-taking, self-confidence, survival attitudes, and field dependence/independence. Scores from these tests were examined for their utility in predicting training outcome (graduation or elimination) and advanced training recommendation (fighter or non-fighter aircraft).

(Continued)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (*Include Area Code*) (512) 536-3877 22c. OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

19. (Concluded)

Results indicated that as a group, the tests demonstrated weak relationships with the performance criteria. No test manifested a consistent pattern of validity for both performance measures. Only the test of self-confidence appeared to contribute to predicting completion of training. Future research efforts are discussed with regard to refining the current test of self-confidence and establishing its construct validity.

AFHRL Technical Paper 87-62

May 1988

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: PRELIMINARY ANALYSIS

Frederick M. Siem
Thomas R. Carretta
Theresa A. Mercatante

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

Lloyd D. Burtch
Chief, Cognitive Skills Assessment Branch

Paper presented at the Third Annual South Texas Symposium on Human Factors and Ergonomics sponsored by the Alamo Chapter of the Human Factors Society, University of Texas at San Antonio, San Antonio, Texas, 8 May 1987.

SUMMARY

A computer-administered test battery, the Basic Attributes Tests (BAT), was administered to 883 USAF pilot candidates. Five tests designed to assess personality and attitudinal characteristics (decisiveness, risk-taking, self-confidence, survival attitudes, and field dependence/independence) were examined for their utility in predicting training outcome (graduation or elimination) and recommendation for advanced training assignment (fighter or non-fighter aircraft).

Results indicated that although the tests demonstrated acceptable reliability, as a group they were not strongly related with the performance criteria. None of the tests was valid against both performance measures. Only the test of self-confidence appeared to contribute to predicting completion of training. Future research efforts are discussed with regard to refining the current test of self-confidence and establishing its construct validity.

PREFACE

This project was conducted under Work Unit 77191845, Selection for Undergraduate Pilot Training, issued by the USAF Air Training Command. This paper was presented at the Third Annual South Texas Symposium on Human Factors and Ergonomics sponsored by the Alamo Chapter of the Human Factors Society. The symposium took place at the University of Texas at San Antonio on 8 May 1987.

This paper is intended to serve as an interim report regarding five tests of personality and attitudinal characteristics from the Basic Attributes Tests (BAT) battery, being considered for use as selection and classification instruments for aircrew candidates.

TABLE OF CONTENTS

LIST OF TABLES

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: PRELIMINARY ANALYSIS

I. INTRODUCTION

Individual differences in student motivation, personality, and attitudes have been acknowledged as important determinants of successful performance in the training environment. The Air Force has a long history of interest in this field of research, particularly with regard to the issue of officer candidate selection (e.g., Flyer & Bigbee, 1955; Guinn, Vitola, & Leisey, 1976; Mullins, 1962; Taylor, Murray, Ellison, & Majesty, 1971; Tupes & Christal, 1957). A major factor contributing to this concern is the high costs associated with candidate attrition, particularly in Undergraduate Pilot Training (UPT). The cost for each UPT eliminee currently is estimated to be about $65,000 to $80,000.

To put current pilot selection research efforts into context, a brief background history follows. Research concerning the selection of military pilot candidates dates back to World War I. Efforts at Kelly Field in 1919 found the highest predictive validity for pilot selection with measures of mental alertness and emotional stability (North & Griffin, 1977). During World War II, the emphasis on pilot selection focused on apparatus tests that were used to measure performance in terms of two-hand coordination, stick-and-rudder skills, rotary pursuit with divided attention, and discrimination reaction time. However, problems with mechanical reliability of the apparatus devices, and a policy shift toward decentralized testing in 1955, led to the discontinuation of apparatus testing in favor of paper-and-pencil tests (Bordelon & Kantor, 1986).

Several research efforts over the next two decades were designed to increase the validity of paper-and-pencil tests for pilot selection through the inclusion of measures of motivation, occupational interests, and personality characteristics (e.g., Guinn et al., 1976; Mullins, 1962; Taylor et al., 1971). Although these studies provided modest support for the use of interest and motivation measures for pilot selection, such tests never were put into operational use.

Technological developments in the 1970s led to a renewed interest in apparatus testing for pilot selection (Hunter & Thompson, 1978; Long & Varney, 1975; McGrevy & Valentine, 1974). The result was development of a computer-administered test battery, the Basic Attributes Tests System, or BAT (Bordelon & Kantor, 1986). The battery includes a number of tests designed to measure characteristics and skills identified as having potential for predicting pilot performance (Imhoff & Levine, 1981). Included among the tests are several tests validated in World War II at the Army Air Forces School of Aviation Medicine, such as measures of psychomotor performance and reaction time (Melton, 1947). Other tests included in the battery are based on more recent advances in the disciplines of cognitive psychology and personality assessment (Imhoff & Levine, 1981). The present focus is on the predictive utility of personality and attitude measures included in the battery.

Five tests from the BAT will be discussed. Two of these tests represent computerized adaptations of paper-and-pencil tests used previously in Air Force research. Dot Estimation, for example, was developed to measure impulsiveness; and Self-Crediting Word Knowledge had been used as a test of self-confidence (Mullins, 1962). Other tests were selected to measure characteristics such as field dependence/independence, risk-taking behavior, and attitudes about risk-taking. These tests were either designed or selected on the basis of their content validity for pilot research. However, it should be noted that the tests selected from the literature had previously been used mainly in academic settings, for purposes other than aircrew selection.

Based on previous research and anecdotal data, it was expected that pilot candidates who successfully completed training would be less impulsive, more self-confident, and more field independent; it also was expected that successful candidates would manifest differences in decision-making style and in attitudes about risk-taking behavior, relative to candidates who failed to graduate from pilot training. Finally, it was expected that within the group of pilots who completed training successfully, there would be differences in characteristics and attitudes between two groups of students: those recommended for advanced training in a fighter-attack-reconnaissance (FAR) track, and those recommended for advanced training in a tanker-transport-bomber track (TTB).

II. METHOD

Subjects

The sample consisted of 883 trainees in the USAF Undergraduate Pilot Training (UPT) program, 867 males and 16 females. In terms of race, the sample consisted of 811 Caucasians, 13 Blacks, and two "Other" (with no data available on 57 subjects). The mean age for the sample was 23.8 years. The composition of the sample in terms of commissioning source is shown in Table 1.

Table 1. Distribution of Subjects by Source (N = 883)

Source	N	
	FSP	Non-FSP
AFROTC Graduate	--	9
AFA Graduate	--	33
USAF Helicopter Pilot	1	--
Other military academy	--	1
OTS Graduate	404	351
USAF Rated Officer	2	1
USAF Non-Rated Officer	6	4
Non-USAF Officer	25	46
Total	438	445

Note. FSP = Flight Screening Program; AFROTC = Air Force Reserve Officer Training Corps; AFA = Air Force Academy; OTS = Officer Training School.

Instrumentation

Embedded Figures

This test has been used to examine the psychological factor of field dependence/independence (Goodenough, 1976; Witkin, 1949). It should be noted that level of field dependence has been treated as a personality characteristic by some researchers and as a perceptual ability by others.

In this test, the subject was presented with a simple geometric figure and two complex figures. The task was to decide which of the two complex figures had the simple figure embedded within it and to indicate a choice by pressing the keypad button corresponding to the figure. Speed and accuracy of response were recorded on each of the 30 trials.

Dot Estimation

This test was designed to study decisiveness (Mullins, 1962). In both the original and the current form of the test, the subject was presented with two boxes, each containing a random display of dots. The subject was instructed to determine, as quickly as possible, which of the two boxes contained more dots. One box always contained one more dot. No instructions were given with regard to counting. The rationale for the development of the test was that decisive individuals would take less time with each item and thus complete more items in a given time period, whereas less decisive individuals would spend more time counting and thus complete fewer items (Mullins, 1962).

In this version of the test, there was a fixed time limit of 5 minutes and 55 items total. Each box was a 3 1/8-inch by 3 1/8-inch square. In the first item, the boxes contained 10 and 11 dots, respectively. With each succeeding item (until the last 15), the number of dots increased by 2 per screen, so that for the last 15 items, the boxes contained 50 and 51 dots, respectively. The measures recorded were the number of trials completed, the number of correct responses, the percent of correct responses, and the total time spent on the test. From these measures was computed the average time spent on correct responses.

Self-Crediting Word Knowledge

This test was designed to be an objective measure of self-confidence (Mullins, 1962). On each of 30 trials, the subject was presented with a "target" word and five other words. The subject's instructions were to choose the synonym closest in meaning to that of the target word.

The test was divided into three sections of 10 questions each. With each succeeding block, the items became more difficult. The subject was informed of this fact and asked to make a bet prior to each of the three blocks. The average number of points bet was recorded, as well as the percent correct, the average reaction time for the correct trials, and a product vector of reaction time and percent correct.

Activities Interest Inventory

This test was developed to measure survival attitudes and risk-taking tendency. The subject was presented with 81 pairs of activities and was asked to indicate a preference for each pair. The subject was told to assume that he/she had the necessary ability to perform each activity. The activity pairs forced the subject to choose between tasks that differed as to threat to

physical survival (sometimes subtly, sometimes not). The measures of interest were the number of high-risk options chosen and the average amount of time required to choose between pairs of activities.

Risk-Taking

This test was selected--on the basis of its previous use in a variety of settings (Slovic, 1966)--to measure risk-taking tendencies when making decisions. Ten boxes were presented in two rows of five boxes each. The subject was told that nine of the boxes contained a reward, and one of the boxes was a "disaster" box. The subject was allowed to select the boxes one at a time. If the selected box contained a payoff, the subject was allowed to accumulate a point for that box. However, if the subject chose the disaster box, all points for that trial were lost.

The test consisted of 30 trials. For 12 of the trials, there was no actual disaster box. The rationale for this design was that on the trials with a disaster box, the "risk trials," the number of boxes selected was impacted by chance selection of the disaster box. On the 12 "no-risk" trials, however, the subject could choose up to nine boxes without actually encountering the disaster box (as the disaster box always appeared as the tenth box). Thus, the no-risk trials were designed to allow subjects to manifest their risk-taking tendencies unaffected by chance selection of an actual disaster box.

The measures recorded on this test were response time and number of boxes chosen separately for the risk and no-risk trials.

UPT Performance Criteria

UPT final training outcome was scored as a dichotomous variable with pass = 1 and fail = 0. Students who passed UPT received a recommendation from an Advanced Training Recommendation Board (ATRB) for advanced training leading to an assignment either as a FAR pilot or a TBB pilot (FAR = 1 and TTB = 0). Final training outcome and ATRB recommendation were determined, in part, by a subject's performance on six check flights during UPT.

Apparatus

The BAT apparatus (Carretta, 1987) consisted of a super-microcomputer built into a self-contained unit with a glare shield and side panels designed to ensure consistency of testing sessions. The subject responded to the various tests using, in combination or individually, a two-axis joystick on the right side of the apparatus, a single-axis joystick on the left side, and a keyboard in the center of the test unit. The keyboard included the numbers 0 to 9, an "Enable" key in the center, and a bottom row with "yes" and "no" keys and two others labeled "S/L" (for same/left responses) and "D/R" (for different/right responses).

Procedure

Most of the subjects in the present sample entered UPT after completion of Officer Training School (OTS). The other subjects came from a variety of other sources (see Table 1). Air Force Reserve Officer Training Corps (AFROTC), OTS, and all active duty USAF candidates were pre-selected on the Air Force Officer Qualifying Test (AFOQT). The AFOQT form used with this sample (Form O; Rogers, Roach, & Wegner, 1986) consisted of 16 subtests which were combined to form five composite measures. All five composite scores were available to OTS selection boards. By Air Force Regulation 51-4, those candidates who were required to take the AFOQT in order to enter UPT had to achieve a minimum score at the twenty-fifth percentile on the Pilot

composite and a minimum score at the tenth percentile on the Navigator-Technical composite, and the sum of the two composite scores had to be equal to or greater than 50.

Prior to entering UPT, about half the sample (n = 438) participated in a 14-hour light aircraft Flight Screening Program (FSP). This program for pilot candidates having no previous flying experience was designed to acquaint them with the aviation environment through flight instruction in a T-41 aircraft. (Candidates such as OTS students with private pilot licenses and Air Force Academy students who completed the standard Academy flight orientation program were exempt from attending FSP.) For the students attending FSP, satisfactory completion of the course was required to continue to UPT.

Subjects were administered the BAT prior to entry into UPT. The test battery consisted of 14 tests that required about 3 1/2 hours to complete. After a test administrator initiated the system, the test session was self-paced by the subject. The test session included programmed breaks between tests to counteract fatigue factors.

All candidates in the sample took part in a UPT program lasting 49 weeks. Candidates could be eliminated from training at any point in the program for a variety of reasons. Those remaining in the program at the forty-second week were considered by an ATRB for either a FAR track or a TTB track.

Analysis

Reliability analyses were conducted for all experimental test measures (on a larger sample from which the present sample was drawn). For the present sample, descriptive statistics were computed, including an intercorrelation matrix of all measures of interest. These measures were examined both for the entire sample and separately for UPT graduates and non-graduates. For UPT graduates only, the data were examined separately for those candidates who were FAR recommended and those who were TTB recommended. A series of regression analyses was conducted. For each regression, predictors were measures from one of the BAT tests, with the criterion either UPT pass/fail or ATRB recommendation.

III. RESULTS

Initial analyses (reported elsewhere; Carretta & Siem, in preparation) were conducted to test for the internal consistency of the measures from four of the five tests. (As Dot Estimation was a speeded test, conventional reliability analyses were not appropriate.) For each of the measures reported, alpha coefficients (Cronbach, 1951) were considered acceptable (greater than .70) with the exception of the measure of the number of boxes chosen on risk trials in the Risk-Taking Test. However, this low reliability was to be expected as there was a chance element determining when a subject would select a disaster box on any particular trial.

Given that the measures appeared to be reliable, the next step was to determine their validity in predicting pilot training outcomes. A series of regression analyses was conducted, two for each of the five tests. One analysis for each test used UPT pass/fail as the dependent measure; for the other, the dependent measure was the ATRB FAR/TTB recommendation.

Descriptive statistics for each measure and their intercorrelations are reported in Tables 2 and 3. The multiple correlation for each test with the two training outcome measures is shown in Table 4. These correlations indicate that the only test to predict UPT pass/fail outcome was the Self-Crediting Word Knowledge Test. Although the correlation was significant (R = .14, F[4,833] = 4.52, p < .01), the magnitude of the relationship was modest. None of the other four tests was

correlated individually with UPT pass/fail, nor were any of the five tests associated with the ATRB recommendation measure.

Table 2. Descriptive Statistics for BAT Test Measures (N = 833)

Time	Mean	SD
Embedded Figures		
Mean RT correct responses (msec)	12,485.76	4,750.92
% correct	64.56	13.86
RTx% correct	2,308.45	68,305.44
Dot Estimation		
# of trials completed	48.72	12.62
# of correct responses	31.31	7.06
Total time spent on test (sec)	147.40	75.27
Mean RT, correct responses (sec)	5.65	4.93
% correct responses	66.23	10.87
Self-Crediting Word Knowledge		
Mean RT, correct responses (msec)	8,214.66	2,055.69
% correct	67.81	10.27
RTx% correct	-3,095.03	30,178.40
Bet	39.02	11.42
Activities Interest Inventory		
# of high-risk choices	50.39	9.74
Mean RT (msec)	4,535.41	1,144.86
Risk-Taking		
# Boxes chosen, risk trials	4.54	.82
Mean RT, risk trials (msec)	2,723.65	1,693.99
# boxes chosen, no-risk trials	6.90	1.33
Mean RT, no-risk trials (msec)	2,277.67	1,577.94

Note. RT x% correct product terms adjusted for mean RT and % correct performance.

The point-biserial correlations with UPT pass/fail for the measures that contributed to the multiple correlation for Self-Crediting Word Knowledge are also shown in Table 4. These correlations indicate that the strongest relationship with training outcome was for the mean response time on trials answered correctly (r [883] = .12, $p < .001$). The percent of correct answers was also significantly but negatively correlated with UPT pass/fail (r [883] = -.07, $p < .05$), as was the amount of points bet (r [883] = -.06, $p < .05$).

Another way to approach understanding these relationships between responses on the Self-Crediting Word Knowledge Test and UPT pass/fail is to examine the means in Table 5. These means indicate that subjects who passed UPT tended to take longer to respond (M = 8.39s vs. M = 7.85s), had fewer correct responses (M = 67.8% vs. M = 68.8%), and bet fewer points on their responses (M = 38.5 vs. M = 40.1).

Table 3. Intercorrelation Matrix of BAT Test Measures (N = 883)

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. EMB Mean RT, correct responses	-																	
2. EMB % correct	-04	-																
3. EMB RTx% correct	-07	-10	-															
4. DOT # of trials completed	-14	-05	02	-														
5. DOT # of correct responses	-09	-01	-04	89	-													
6. DOT Total time spent on test	20	08	-02	-76	-61	-												
7. DOT Mean RT, correct responses	18	06	-01	-93	-85	88	-											
8. DOT % correct responses	13	10	-09	-70	-34	62	62	-										
9. WKA Mean RT, correct responses	15	-03	-01	-16	-12	18	16	14	-									
10. WKA % correct	-07	08	03	06	-06	-05	-07	-02	-15	-								
11. WKA RTx% correct	00	08	01	01	02	-01	-02	01	-20	-17	-							
12. WKA Bet	-10	10	-03	-06	-05	-01	03	04	-25	27	04	-						
13. AIA # of high-risk choices	-10	05	-03	-04	-03	-02	01	-03	-06	10	04	16	-					
14. AIA Mean RT	10	11	-06	-11	-05	15	12	16	44	-13	-03	-12	-11	-				
15. RSK # of boxes, risk trials	06	-04	02	06	07	-06	-08	-02	-03	02	02	01	04	-08	-			
16. RSK Mean RT, risk trials	04	01	-04	04	03	03	-01	-03	11	-03	-01	-22	-06	11	.05	-		
17. RSK # of boxes, no-risk trials	04	-09	00	12	11	-11	-13	-08	-08	01	04	-03	06	-11	59	01	-	
18. RSK Mean RT no-risk trials	03	00	-04	03	02	03	00	-02	10	-04	00	-24	-06	11	-05	96	01	-

Note. EMB = Embedded Figures; DOT = Dot Estimation; WKA = Self-Credibing Word Knowledge; AIA = Activities Interest Inventory; RSK = Risk-Taking.

Table 4. Correlational Relationships Between BAT Tests and Training Outcomes

Test	UPT pass/fail (N = 883)	FAR/TTB (N = 519)
Embedded Figures		
Mean RT, correct responses	.00	-.03
% correct	-.06	.05
RTx% correct	.00	.04
Multiple Correlation	.06	.08
Dot Estimation		
# of trials completed	.00	-.05
# of correct responses	.01	.00
Total time spent on test	.02	.03
Mean RT, correct responses	.00	.05
% correct responses	.01	.07
Multiple Correlation	.03	.14
Self-Crediting Word Knowledge		
Mean RT, correct responses	.12***	-.05
% correct	-.07*	-.03
RTx% correct	.03	.06
Bet	-.06*	.02
Multiple Correlation	.14**	.08
Activities Interest Inventory		
# of high-risk choices	-.02	.04
Mean RT	-.06	-.05
Multiple Correlation	.06	.07
Risk-Taking		
# boxes chosen, risk trials	-.03	.02
Mean RT, risk trials	-.04	-.01
# boxes chosen, no-risk trials	-.03	.01
Mean RT, no-risk trials	-.03	-.02
Multiple Correlation	.05	.06

Note. RT = Response Time; # = Number; % = Percent; RTx% = vector product term. Significance levels for zero-order correlations reported only when multiple correlation is significant.

*p < .05. **p < .01. ***p < .001.

Table 5. Mean Scores for Self-Crediting Word Knowledge by Training Outcome

Test measure		UPT fail (N = 284)	UPT pass (N = 599)
RT (msec)	M	7,854.38	8,385.48
	SD	1,753.79	2,164.79
% correct	M	68.84	67.81
	SD	10.03	10.35
Bet	M	40.08	38.51
	SD	10.25	11.91

Note. Response time (RT) in milliseconds.

For a subset of the sample who had data on the Word Knowledge subtest of the AFOQT, additional analyses were conducted to examine differences on the Self-Crediting Word Knowledge items, controlling for overall verbal ability. The pattern of adjusted means was similar to that for the unadjusted means (see Table 6), although the differences were smaller. Note, however, that graduates and eliminees differed significantly in average reaction time, even after controlling for verbal ability. These data tend to suggest that candidates who successfully completed UPT tended to take longer in their responses, which could be interpreted as a manifestation of a more cautious decision-making style.

Table 6. Mean Score for Self-Crediting Word Knowledge by Training Outcome Adjusted for AFOQT Word Knowledge Scores

Measure	UPT fail (N = 259)	UPT pass (N = 488)
RT (msec)	7,919.22	8,429.67
% correct	68.11	68.01
Bet	39.42	39.07

Additional evidence for this interpretation of the data comes from comparing UPT graduates with non-grads in terms of their responses on measures from the other tests (although these differences were not statistically significant). Candidates who passed UPT selected, on the average, fewer high-risk choices on the Activities Interest Inventory ($\underline{M}$ = 50.25 vs. $\underline{M}$ = 50.70 [out of 81 items]). UPT graduates also were "less impulsive" on the Dot Estimation Test; that is, relative to UPT failures, the successful students completed fewer of the self-paced trials ($\underline{M}$ = 48.71 vs. $\underline{M}$ = 48.89), spent more time on the test ($\underline{M}$ = 148.26s vs. $\underline{M}$ = 146.11s), and spent more time on correct responses ($\underline{M}$ = 56.62 vs. $\underline{M}$ = 56.42). The UPT graduates also answered more of the items correctly ($\underline{M}$ = 66.33% vs. $\underline{M}$ = 66.25%); although the differences on the Dot Estimation measures between UPT graduates and eliminees were not statistically significant, they are consistent with the interpretation of responses to the Self-Crediting Word Knowledge Test that UPT graduates were more cautious and deliberate, less impulsive and hasty, in their approaches to the experimental tests than were the UPT failures.

IV. DISCUSSION

Each of the five BAT tests exhibited acceptable reliability for use as selection devices. However, none of them was related statistically to both measures of training performance (graduation/elimination and advanced training assignment). There are several explanations for the poor predictive utility demonstrated by these tests. One explanation is that the tests may not have been measuring the characteristics that they were designed to measure (i.e., poor construct validity). Although all but one of the tests was adapted from a previously validated paper-and-pencil test, no subjects were given both the original form and the computerized version of the same test. That is, some facet of computer administration of the test may have changed the characteristics of the test.

Another possible explanation for the low validity coefficients for these tests is a restriction in the range of subject variability, in terms of test performance. That is, the

candidates who entered UPT may have been a fairly homogeneous group, as most already had been screened on the AFOQT and some form of flight training. In some fashion, those "gates" may have selected out candidates who differed from those actually entering UPT, in terms of risk-taking behaviors and attitudes, self-confidence, and the other characteristics that these tests were designed to measure.

Finally, these tests may be accurately assessing what they purport to measure, but those characteristics may not be relevant to predicting success in pilot training. Even with the one test demonstrating modest validity in the present sample, Self-Crediting Word Knowledge, the average performance difference between passes and non-graduates was fairly small (see Table 6). However, in analyses reported elsewhere (Carretta, 1987), this test continued to be a significant predictor of UPT performance when included in a prediction model that consisted of a number of psychomotor and cognitive skills tests, the types of tests that have been used more traditionally for pilot selection. Thus, it appears that the Self-Crediting Word Knowledge Test may contribute unique predictive validity to a selection system for pilot candidates.

V. CONCLUSION

Future efforts will be directed at improving and refining the present version of the Self-Crediting Word Knowledge Test. Other research efforts currently underway are designed to determine the test's construct validity. In order to develop a better idea of exactly what the test is measuring, more traditional personality tests of characteristics such as self-confidence (e.g., Spence, Helmreich, & Holahan, 1979) are being administered to Air Force pilots with varying levels of training experience who have been tested previously on the BAT. By examining relationships between the traditional, validated tests of self-confidence with the Self-Crediting Word Knowledge measures, a better understanding of the characteristics of the test and how it can be implemented in future models for Air Force pilot selection and classification can be developed.

REFERENCES

Air Force Regulation 51-4. (1983, July). Application procedures for Undergraduate Flying Training (UNT). Washington, DC: Department of the Air Force.

Bordelon, V.P., & Kantor, J.E. (1986, July). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (AFHRL-TR-86-4, AD-A170 353). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987, April) The Basic Attributes Tests: An experimental selection and classification instrument for U.S. Air Force pilot candidates. Paper presented at the Fourth International Symposium on Aviation Psychology, Columbus, Ohio.

Carretta, T.R., & Siem, F.M. (in preparation). Personality, attitudes, and pilot training performance: Final analyses. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cronbach, L.J. (1951). Coefficient alpha and the iternal structure of tests. Psychometrika, 16, 297-334.

Flyer, E.S., & Bigbee, L.R. (1955). Validity of AFOQT aptitude and interest scores as predictors of AFROTC pilot training success (PRL-TM-55-35). Lackland AFB, TX: Personnel Research Laboratory, Air Force Personnel and Training Research Center.

Goodenough, D.R. (1976). The role of individual differences in field dependence as a factor in learning and memory. Psychological Bulletin, 83, 675-694.

Guinn, N., Vitola, B.M., & Leisey, S.A. (1976). Background and interest measures as predictors of success in undergraduate pilot training (AFHRL-TR-76-9, AD-A025 851). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hunter, D.R., & Thompson, N.A. (1978). Pilot selection system development (AFHRL-TR-78-33, AD-A054 418). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Imhoff, D.L., & Levine, J.M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHRL-TR-80-27, AD-A094 317). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Long, G.E., & Varney, N.C. (1975). Automated pilot aptitude measurement system (AFHRL-TR-75-58, AD-A018 151). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

McGrevy, D.F., & Valentine, L.D., Jr. (1974). Validation of two aircrew psychomotor tests (AFHRL-TR-74-4, AD-777 830). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Melton, A.W. (Ed). (1947). Apparatus tests (Air Force Aviation Psychology Program Research Report No. 4). Washington, DC: U.S. Government Printing Office.

Mullins, C.J. (1962). Objective tests of self-confidence (PRL-TM-62-6). Lackland AFB, TX: Selection and Classification Branch, Personnel Research Laboratory.

North, R., & Griffin, G. (1977). Aviation selection 1919-1977 (Special Report SR77-2). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Rogers, D.L., Roach, B.W., & Wegner, T.G. (1986). Air Force Officer Qualifying Test Form O: Development and standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Slovic, P. (1966). Risk-taking in children: Age and sex differences. Child Development, 37, 169-176.

Spence, J.T., Helmreich, R.L., & Holahan, C.K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. Journal of Personality and Social Psychology, 37, 1673-1682.

Taylor, C.W., Murray, S.L., Ellison, R.L., & Majesty, M.S. (1971, July). Development of motivation assessment techniques for Air Force officer training and education programs: Motivation for pilot training (AFHRL-TR-71-21, AD-751 487). Brooks AFB, TX: Professional Education Division, Air Force Human Resources Laboratory.

Tupes, E.C., & Christal, R.E. (1957, April). Psychological tests and the selection and classification of Air Force officers (AFPTRC-TN-57-52, AD-126 383). Lackland AFB, TX: Personnel Laboratory, Air Force Personnel and Training Research Center.

Witkin, H.A. (1949). Perception of body position and the position of the visual field. Psychology Monographs, 63, 1-46.